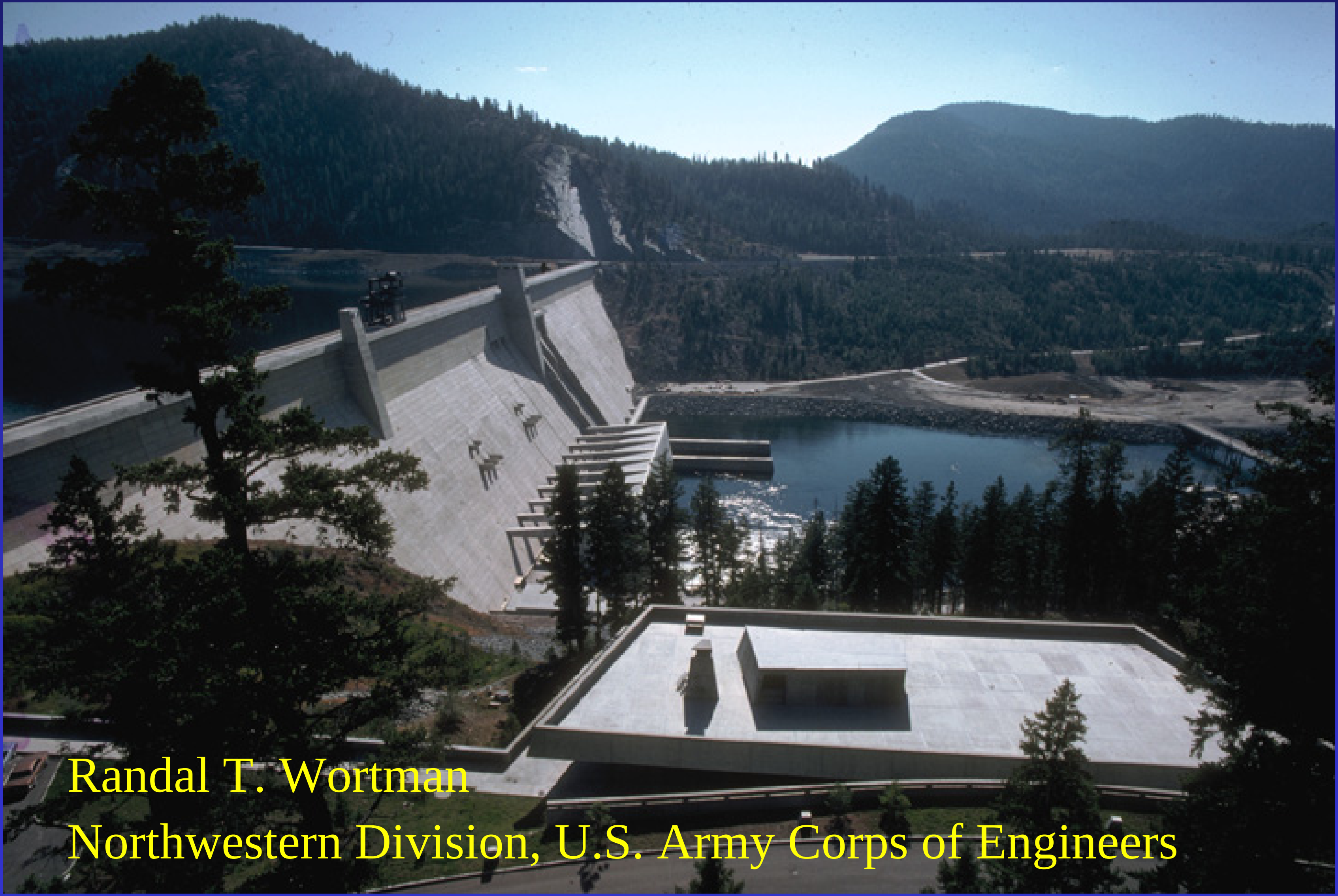
US Army Corps
of Engineers
Northwestern Division

Regression Techniques in Water Supply Forecasting



Randal T. Wortman
Northwestern Division, U.S. Army Corps of Engineers

US Army Corps
of Engineers
Northwestern Division

Where we are headed…

Observed Data

$X1_1$	$X2_1$	$X3_1$
$X1_2$	$X2_2$	$X3_2$
$X1_3$	$X2_3$	$X3_3$
$X1_4$	$X2_4$	$X3_4$
$X1_5$	$X2_5$	$X3_5$
$X1_6$	$X2_6$	$X3_6$
…	…	…
$X1_N$	$X2_N$	$X3_N$

Traditional Regression Model

$$Y = \beta_0 + \beta_1 * X1 + \beta_2 * X2 + \beta_3 * X3$$

Principal Components

$PC1_1$	$PC2_1$	$PC3_1$
$PC1_2$	$PC2_2$	$PC3_2$
$PC1_3$	$PC2_3$	$PC3_3$
$PC1_4$	$PC2_4$	$PC3_4$
$PC1_5$	$PC2_5$	$PC3_5$
$PC1_6$	$PC2_6$	$PC3_6$
…	…	…
$PC1_N$	$PC2_N$	$PC3_N$

Principal Component Regression Model

$$Y = \beta_0 + \beta_1 * PC1 + \beta_2 * PC2 + \beta_3 * PC3$$

US Army Corps
of Engineers
Northwestern Division

Standard Multiple-Variable Regression

- *Dependent variable is a seasonal inflow volume*
- *Predictor variables are pseudo-variables created from weighted combinations of similar stations (e.g. sum or average of three snow stations)*
- *Some models include variables not yet observed ("future variables") where the historic mean ("normal subsequent") is substituted in lieu of the variable being measured, e.g. using "May+June precip" in a 1-April forecast*

$$Y = \beta_0 + \beta_1 * X1 + \beta_2 * X2 + \beta_3 * X3 \ldots$$

US Army Corps
of Engineers
Northwestern Division

Regression Equation for Libby Water Supply Forecast

$$LibbyLocal_{APRAUG} = \beta_0 + \beta_1 * \sum_{Oct}^{Nov} LibbyLocal\ Runoff$$

$$+ \beta_2 * \sum_{Oct}^{Mar} (ELKB + FENB + FTIM + LRSM + BFEI + POLM)WinterPrecip$$

$$+ \beta_3 * (SUMB + NFRB + RMTM + KIMB + WSLM + 0.5 * MORB)Apr1\_Snow$$

$$+ \beta_4 * \sum_{Apr}^{Aug} w_i (FTIM + PTHI + KASB + WHFM)SpringPrecip$$

$$w_i \in \{1.0, 1.0, 0.8, 0.5, 0.2\}$$

US Army Corps
of Engineers
Northwestern Division

Concerns with Traditional Regression Models

- *Subjective station selection*
- *Subjective station weighting/aggregating*
- *Use of "normal subsequent" variable as a surrogate for a "future value" variable*
- *Variable selection and model selection based on model calibration statistics*
- *Predictor variables frequently highly intercorrelated*

US Army Corps
of Engineers
Northwestern Division

Principal Components

- *Creates surrogate variables (principal components) that are a weighted combination of the original variables.*
- *The principal components are fully independent of each other (zero intercorrelation)*
- *Most of the “variability” is loaded into the first one or two components.*
- *Eigenvalues reflect the proportion of the variability in original variables loaded into each component*
- *Note: PCs combine the information <u>within the predictor variables</u>, but “have no knowledge” of the dependent variable, the variable to be forecasted.*

Principal Components

Observed Data			Principal Components		
$X1_1$	$X2_1$	$X3_1$	$PC1_1$	$PC2_1$	$PC3_1$
$X1_2$	$X2_2$	$X3_2$	$PC1_2$	$PC2_2$	$PC3_2$
$X1_3$	$X2_3$	$X3_3$	$PC1_3$	$PC2_3$	$PC3_3$
$X1_4$	$X2_4$	$X3_4$	$PC1_4$	$PC2_4$	$PC3_4$
$X1_5$	$X2_5$	$X3_5$	$PC1_5$	$PC2_5$	$PC3_5$
$X1_6$	$X2_6$	$X3_6$	$PC1_6$	$PC2_6$	$PC3_6$
…	…	…	…	…	…
$X1_N$	$X2_N$	$X3_N$	$PC1_N$	$PC2_N$	$PC3_N$

US Army Corps
of Engineers
Northwestern Division

Principal Components Regression

- ***Methodology:***
  - P-variables * weighting matrix -> P principal components
  - Multivariable regression:
    - Use the P principal components as the predictor variables
    - Use the original WSF runoff variable as the dependent variable.

$$Y = \beta_0 + \beta_1 * PC1 + \beta_2 * PC2 + \beta_3 * PC3 \ldots$$

US Army Corps
of Engineers
Northwestern Division

Principal Components Regression

- ***Properties of the Principal Component Regression Model with all "P" Components ("p"= # of original variables)***
  - Component R-squared values
  - Component R-squared loading
  - Eigenvalue loading
- **7 original variables -> 7 principal components:**

PC Analysis	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7
Component R-Square	0.84144	0.00383	0.00040	0.00003	0.00495	0.00336	0.00068
Component R-Square %	98.5%	0.4%	0.0%	0.0%	0.6%	0.4%	0.1%
Cumulative R-Square	0.8414	0.8453	0.8457	0.8457	0.8506	0.8540	0.8547

Eigenvalues	4.41	0.80	0.74	0.55	0.36	0.09	0.04

Principal Components Regression

- ***Variable Selection (Which components do I keep?)***
  - **Component retention criteria:**
    - **Eigenvalue: provides the proportion of variability of X variables contained in each PC**
    - **significant R-squared: indicates the variability in the Y variable explained by this PC in a linear model, i.e. the usefulness of this PC in predicting the Y variable**
    - **significant *Beta:* reject this component when the regression coefficient is indistinguishable from zero.**
    - **sign of *Beta*: be wary of this component if the sign is negative (applies to water supply forecasting)**

US Army Corps
of Engineers
Northwestern Division

Dworshak Water Supply Forecast Example

- ***Predictor variables for Dworshak Water Supply Forecast:***
  - ***Climate variable: September SOI***
  - ***Precip variables: Monthly precip at Elk River for Dec & Jan***
  - ***Snow variables: 1-March SWE at 4 locations***
- ***We will be looking at the following:***
  - ***Correlation Analysis – Original Variables***
  - ***Standard Regression***
  - ***Stepwise Regression***
  - ***Correlation Analysis – Principal Components***
  - ***Principal Components Regression (All components)***
  - ***Principal Components Regression (Selected components)***

US Army Corps
of Engineers
Northwestern Division

Dworshak Water Supply Forecast Example

- ***Correlation Analysis – Original Variables (7)***

	Dwr_AprJul	*SOI_Sep*	*ELKDecPpt*	*ELKJanPpt*	*PierceMar*	*LostLkMar*	*HoodooMar*	*ElkButteMar*	*WYear*
Dwr_AprJul	1								
SOI_Sep	0.6309	1							
ELKDecPpt	0.5039	0.2665	1						
ELKJanPpt	0.5953	0.2872	0.2700	1					
PierceMar	0.7413	0.5857	0.3648	0.4760	1				
LostLkMar	**0.8367**	0.4140	0.4886	0.5128	0.6744	1			
HoodooMar	**0.8441**	0.4811	0.4939	0.5712	0.6276	**0.9300**	1		
ElkButteMar	**0.8664**	0.4798	0.4401	0.5696	**0.8010**	**0.9078**	**0.8495**	1	
WYear	-0.1818	-0.1211	-0.0647	-0.1697	-0.1183	-0.3165	-0.3073	-0.1903	1

US Army Corps
of Engineers
Northwestern Division

Dworshak Water Supply Forecast Example

- ***Standard Regression (fits all 7 predictor variables)***

Regression Statistics	
Multiple R	0.924488
R Square	0.854679
Adjusted R Square	0.826422
Standard Error	373.218
Observations	44

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	401.961	239.526	**1.678**	**0.102**
SOI_Sep	202.159	67.586	2.991	0.005
ELKDecPpt	33.610	27.035	**1.243**	**0.222**
ELKJanPpt	44.749	30.768	**1.454**	**0.155**
PierceMar	5.537	25.381	**0.218**	**0.829**
LostLkMar	6.504	13.613	**0.478**	**0.636**
HoodooMar	14.519	15.053	**0.964**	**0.341**
ElkButteMar	25.593	13.992	1.829	0.076

US Army Corps
of Engineers
Northwestern Division

Dworshak Water Supply Forecast Example

- ***Stepwise Regression (fits 3 predictor variables)***

Regression Statistics	
Multiple R	0.916011
R Square	0.839077
Adjusted R Square	0.827008
Standard Error	372.5874
Observations	44

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	508.027	222.764	2.281	0.028
SOI_Sep	201.763	59.391	3.397	0.002
HoodooMar	25.021	9.449	2.648	0.012
ElkButteMar	33.801	8.689	3.890	0.000

US Army Corps
of Engineers
Northwestern Division

Dworshak Water Supply Forecast Example

- ***Correlation Analysis – Principal Components***

	Dwr_AprJul	*PC7_1*	*PC7_2*	*PC7_3*	*PC7_4*	*PC7_5*	*PC7_6*	*PC7_7*	*WYear*
Dwr_AprJul	1								
PC7_1	**0.9173**	1							
PC7_2	-0.0619	0.0000	1						
PC7_3	-0.0199	0.0000	0.0000	1					
PC7_4	-0.0054	0.0000	0.0000	0.0000	1				
PC7_5	0.0703	0.0000	0.0000	0.0000	0.0000	1			
PC7_6	0.0579	0.0000	0.0000	0.0000	0.0000	0.0000	1		
PC7_7	0.0260	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1	
WYear	-0.1818	-0.2436	-0.0389	-0.0909	-0.0898	-0.2064	0.1400	0.1902	1

US Army Corps of Engineers
Northwestern Division

Dworshak Water Supply Forecast Example

- ***Principal Components Regression (All 7 components)***

PC regression with all components is identical to standard regression with all variables =>

Regression Statistics	
Multiple R	0.924489
R Square	0.854679
Adjusted R Square	0.826422
Standard Error	373.218
Observations	44

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2627.557	56.265	46.700	0.000
PC_1	391.149	27.092	14.438	0.000
PC_2	-61.982	63.613	**-0.974**	**0.336**
PC_3	-20.734	66.204	**-0.313**	**0.756**
PC_4	-6.565	76.988	**-0.085**	**0.933**
PC_5	104.582	94.466	**1.107**	**0.276**
PC_6	170.545	187.024	**0.912**	**0.368**
PC_7	110.065	268.496	**0.410**	**0.684**

US Army Corps
of Engineers
Northwestern Division

Dworshak Water Supply Forecast Example

- Select Components to Keep (or use Stepwise on PCs)
  - Component R-squared values
  - Component R-squared loading
  - Eigenvalue loading

PC Analysis	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7
Component R-Square	0.84144	0.00383	0.00040	0.00003	0.00495	0.00336	0.00068
Component R-Square %	98.5%	0.4%	0.0%	0.0%	0.6%	0.4%	0.1%
Cumulative R-Square	0.8414	0.8453	0.8457	0.8457	0.8506	0.8540	0.8547

Eigenvalues	4.41	0.80	0.74	0.55	0.36	0.09	0.04

US Army Corps
of Engineers
Northwestern Division

Dworshak Water Supply Forecast Example

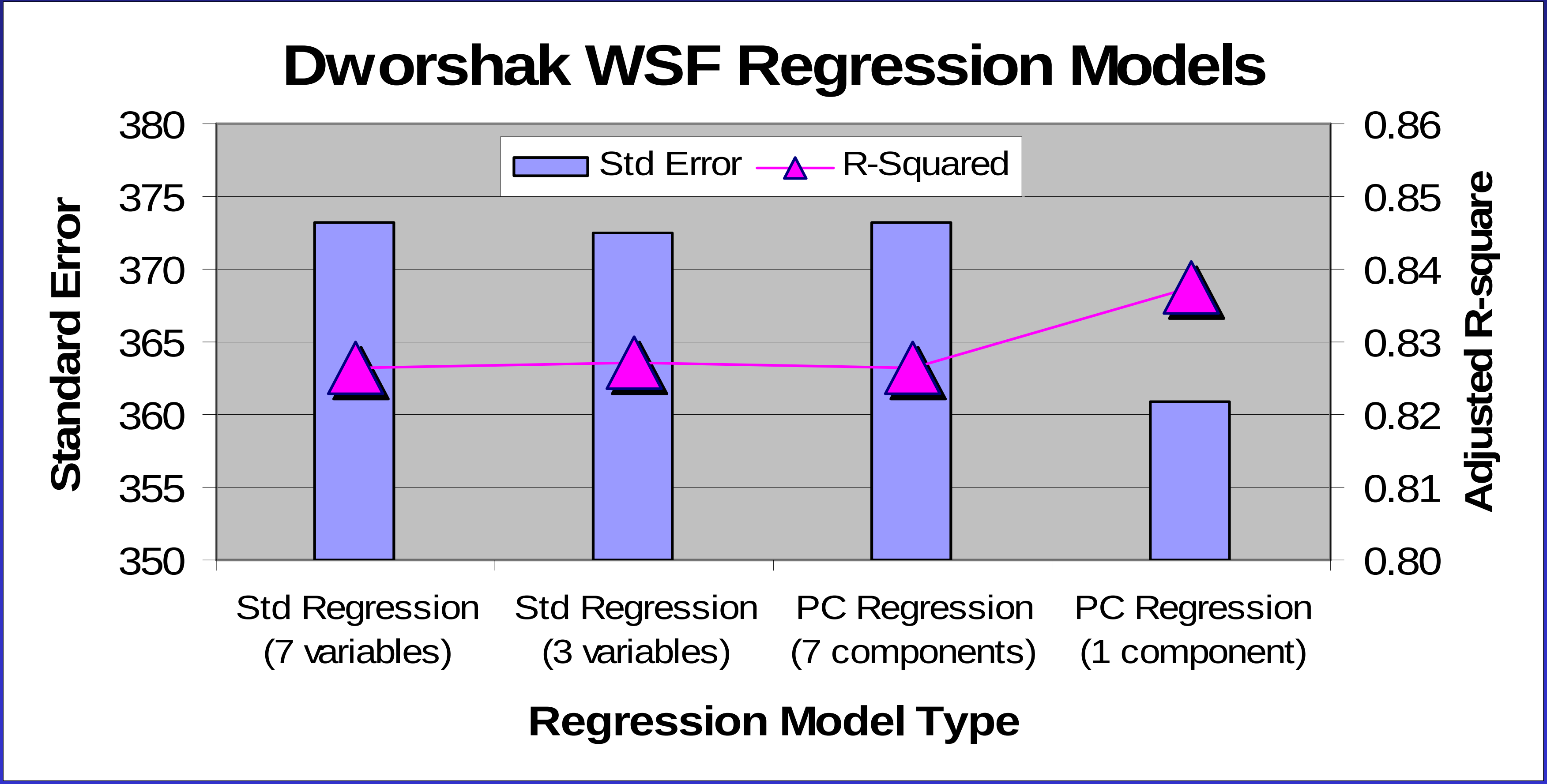
- ***Principal Components Regression (1 component)***

Regression Statistics	
Multiple R	0.917301
R Square	0.841440
Adjusted R Square	0.837665
Standard Error	360.928
Observations	44

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2627.561	54.412	48.290	0.000
PC7_1	391.149	26.200	14.929	0.000

US Army Corps
of Engineers
Northwestern Division

Model Comparison:
Standard Error & R-Squared statistics



US Army Corps
of Engineers
Northwestern Division

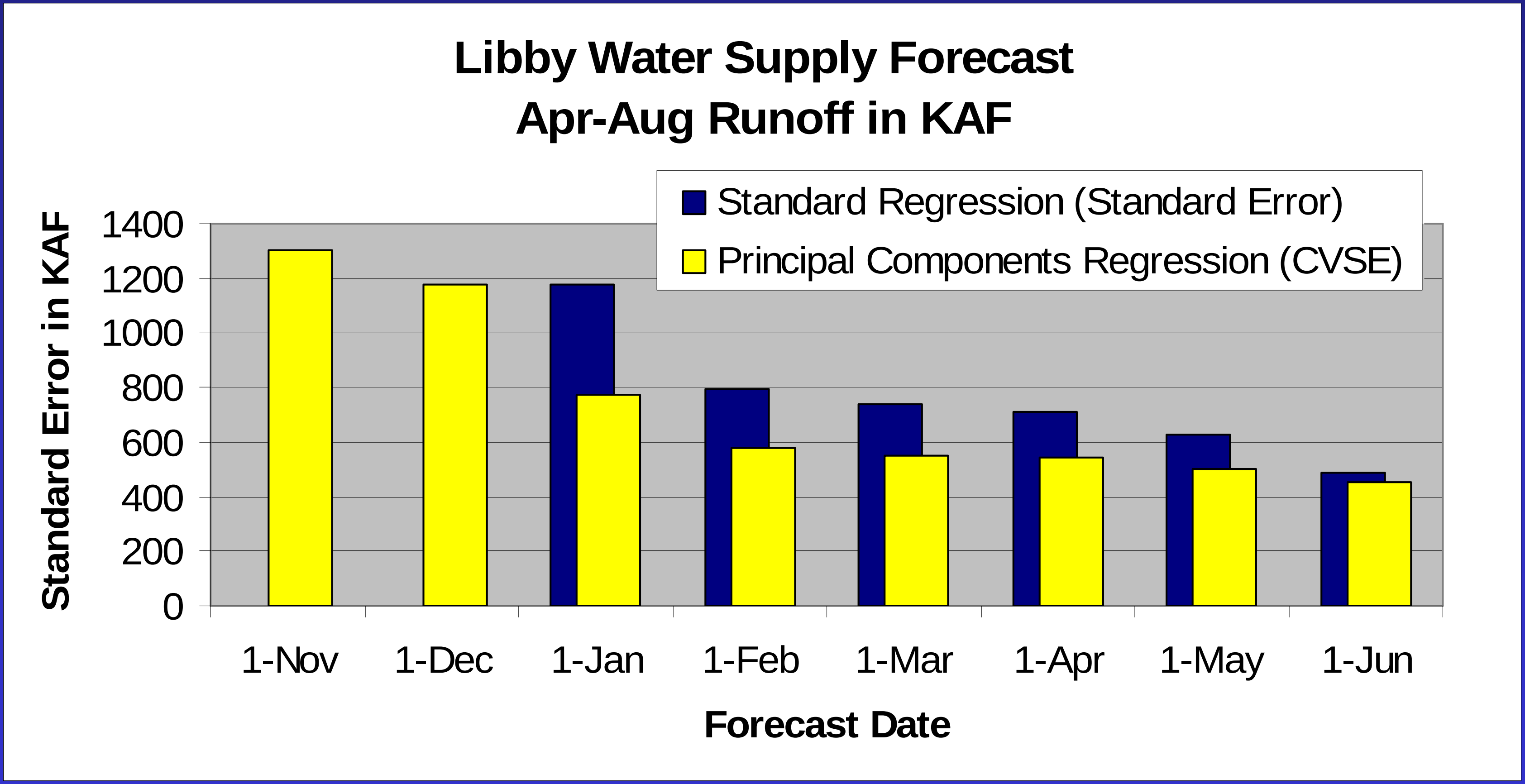
Model Comparison:
Cross-Validation Standard Error (CVSE)

- *Calibration vs. Validation statistics*
- *Split-Sample validation*
- *Leave-one-out validation*
- *Cross-validation standard error (“Jackknife” Std Err)*
- *CVSE supports model parsimony by including d-f adj.*
- *CVSE better indicator of how the model performs with data not used in the calibration, that it “hasn’t seen yet”*
- *CVSE = PRESS statistic adjusted for degrees-of-freedom*
- *CVSE can be directly calculated from either Projection matrix or “Hat” matrix (calculated by NRCS PCREG)*

US Army Corps of Engineers
Northwestern Division

Libby Water Supply Forecast using Principal Components Regression

- *New equation for each month (similar pool of variables)*
- *Standard Error/CVSE plots*



Software for Regression Analysis

Capability	Software Package		
	Statgraphics	MS Excel	NRCS PCREG
Standard Regression	X	X	X
Stepwise Regression	X	X	
Adj R-square	X	X	
Standard Error	X	X	X
Principal Components	X	add-in	X
Eigen analysis	X	add-in	X
PrinComp Regression	X	indirectly	X
Cross Validation SE		indirectly	X
Optimal Search			X
Graphics	X	X	

US Army Corps
of Engineers
Northwestern Division

Endorsements

The following agencies use Principal Components regression in their Water Supply Forecasting procedures:

- *National Water and Climate Center, Natural Resources Conservation Service*
- *Northwestern Division, U.S. Army Corps of Engineers*
- *Northwest River Forecasting Center, National Oceanic and Atmospheric Administration*
- *Columbia River Treaty Operating Committee; Canadian and United States Entities*
- *Bonneville Power Administration, Dept of Energy*
- *BC Hydro*

Questions?